LLM-informed drone visual inspection for civil infrastructure

Jiucai Liu, Chengzhang Chai, Haijiang Li*, Yan Gao, Xiaofeng Zhu
BIM for Smart Engineering Centre, School of Engineering, Cardiff University, UK
LiuJ151@cardiff.ac.uk

Abstract. Drone-based visual inspection has emerged as a crucial manner for infrastructure inspection, owing to its mobility and potential for automated perception. However, from the perspective of human-agent interaction, the predominant modes in this field currently involve either full-process human intervention or end-to-end execution based on deep learning. The former constrains the level of automation, while the latter overlooks considerations of interactivity and controllability. To improve the level of automation and interactivity of the inspection process, this research explores the integration of Large Language Models (LLMs) into the visual inspection of infrastructure to utilize the capabilities of LLMs to understand human intentions and generate control commands. Specifically, high-level function libraries for drone and sensor control, as well as comprehensive and standardized prompts, are developed to fulfill the objective. The effectiveness of the method is demonstrated in both simulated and laboratory environments.

1. Introduction

The increasing pressure on infrastructures due to higher demands for recycling and resource efficiency (Parliament, E.U., 2011), alongside the deterioration of thousands of in-service infrastructures from issues like corrosion and cracks, highlights the growing necessity for long-term maintenance with higher efficiency and automation. Visual inspection is essential for evaluating infrastructure conditions and drone-based visual inspection has shown promising automation and standardization in practice (Nooralishahi et al., 2021) compared with traditional manual inspection.

Specifically, drone-based visual inspection for infrastructure has shown promising effects on leveraging the mobility of drones (Nooralishahi et al., 2021), the perception ability of onboard sensors, and the analytical ability of artificial intelligence methods. Many studies tend to concentrate on specific technologies in the drone-based visual inspection process, overlooking the working patterns involved. While considering human-agent interaction, the majority of research on drone-based visual inspection either leans heavily on human control with minimal automation or entirely relies on end-to-end deep learning methods to perform tasks without human intervention, which poses a risk of low data quality leading to recapture in the aftermath.

In this paper, to further improve the automation and interactiveness of drone-based visual inspection for infrastructure and ensures a thorough and systematic inspection process, large language model (LLM) was introduced in the loop of process. Recent advancements in LLM and its exceptional performance in text generation, machine translation, and code synthesis have make it possible for this development.

Several steps are taken to demonstrate the potential of LLM for drone-based visual inspection in infrastructure. Firstly, we introduce LLM into the process of drone-based visual inspection and design the scheme for LLM-informed drone-based visual inspection for infrastructure. Secondly, high-level function libraries about drone control, sensor control and algorithm API, which act as a bridge from human intent and control command to restrict the output of LLM, are created to empower the LLM model to generate code for drone navigation, data acquisition and data post-processing. Thirdly, information about high-level function libraries and the proposed inspection scheme was translated into several prompts to provide essential information for LLM. The interacting pattern between LLM and engineers is also defined to make the process more understandable, more explicit, and more controllable. Finally, to validate the effectiveness and adaptability of our method, we conducted visual inspection experiments in a simulated environment containing a bridge and some other buildings. Then, we transform our system into a physical drone to conduct experiments in a lab environment.

The main contributions of this paper are listed below:

1. Introduce LLM into drone-based visual inspections for infrastructure to enhance the level of automation of information collection by providing a comprehensive and interactive approach. As an embodied agent, LLM can assist in task planning by understanding human intentions and adapting the plans based on human interactions. Moreover, LLM can collaborate with tools like drone control algorithms, sensor control algorithms, and other AI models for data interpretation, ensuring the successful completion of the inspection objectives.

2. Develop a set of high-level function libraries to empower LLM to generate standardized control commands for drone control, data collection and data post-processing.

3. Develop the systematic prompts introducing the proposed inspection scheme, high-level function libraries and engineer-LLM interacting pattern to fulfil the LLM-informed drone visual inspection system.

4. Validate the effectiveness of the proposed method in both the simulated environment and the lab environment.

2. Related work

In this section, the discussion about the related literature on drone-based visual inspection for infrastructure and large language models in robotics will be proposed.

2.1 Drone-based visual inspection for infrastructure

Drone-based visual inspection has proven to offer several advantages over traditional inspection methods, including cost, time, reduced risk for inspectors, and inspection quality (Chan et al., 2015). Additionally, drones with high mobility and suitable sensing systems can capture various types of data, such as RGB images, cloud points, and thermal images, for different objectives (Zhang et al., 2022). Due to these advantages, drones have been widely used for the non-destructive inspection of infrastructure (Cheng et al., 2020). Besada et al. (2018) proposed the mission definition system of drone-based visual inspection for specific infrastructure and split the whole task into flight plan calculation, trajectory calculation, flight generation and measurement translation. Jung et al. (2018) introduced a practical 3D coverage path planning method for drone-based inspection of high-rise structures by dividing the predefined volumetric map of structures into several layers to improve the resolution of the data collection.

However, to fully fulfil the automatic visual inspection, it is crucial to incorporate human intervention into the inspection process to adapt the inspection strategy promptly. When exploring human-agent interaction, the prevailing research on drone-based visual inspection tends to either heavily rely on human control with minimal automation or solely rely on end-to-end deep learning methods to execute tasks without human intervention posing a significant risk of compromised data quality, ultimately necessitating additional capture efforts afterwards.

The blooming development of LLM, especially ChatGPT has make it possible to enable engineers to participate in the dynamic decision-making process of visual inspection and further enhance interactivity and automation of the process.

2.2 LLM-based agent: applying LLM in industrial scenario

A large language model is a neural network with hundreds of billions of parameters trained on unsupervised learning objectives such as next-token prediction or masked-language modelling. It exhibits remarkable multi-task generalization, promising capability of language generation (including programming) and interactivity, which enable it to interact with users and to leverage various objects, including robotics, sensors, algorithms, and even other AI models. The versatility and scalability have sparked growing research interest in developing LLM-based agents for industrial applications.

Some researchers aim to harness LLM's reasoning capability to generate feasible plans for complex scenarios and fulfil automatic and interactive robotic control objectives. Ahn et al. (2022) proposed a method called SayCan which enables the leveraging of the rich knowledge in pre-trained large language models to complete embodied tasks. They also demonstrated that the robot's performance can be improved simply by enhancing the underlying language model. On the basis of SayCan framework, Chen et al. (2022) further introduced a flexible and quarriable spatial semantic representation called NLMap based on visual-language models, including ViLD and CLIP, and significantly improved long-horizon planning via natural language instructions in the open-world domain,

Apart from interacting with sensors and robots, LLM is also used to interact with other AI models to form a systematic AI agent in complex scenarios. Shah et al. (2022) presented LM-Nav, a system combined with LLM, Visual Language Navigation (VLN) and Visual Navigation Model (VNM), which can navigate robot from textual instructions without requiring any user annotations for navigational data. In the system, the LLM is responsible for parsing user instructions into a list of landmarks, the VLN is responsible for estimating the probability that each observation in a "mental map" constructed from prior exploration of the environment matches these landmarks and the VNM is responsible for estimating navigational affordances (distances between landmarks) and robot actions.

Research has also been done on the application of LLM for drone control recently. Lamine et al. (2023) introduced LLM in the natural language-based drone control systems and proved its usability and reliability. But their research only focusses on the drone control without concerning context information of the actual task. Our research will make the first step for applying the LLM-informed system in the application of visual inspection for infrastructure.

3. Methodology

To enhance the level of automation and interactivity of current drone-based visual inspection methods and elevate the level of automation and interactiveness in the inspection process, a holistic framework for LLM-informed visual inspection of infrastructure is put forth. This framework incorporates LLM to streamline and optimize the inspection procedures adaptively based on human intention.

3.1 LLM-informed visual inspection framework for infrastructure

In the paradigm shown in Figure 1, we introduce LLM as an agent into the task of visual inspection for infrastructure. LLM as agents are capable of observing, acting, and receiving feedback iteratively from external entities including drones, sensors algorithms and other AI models (Wang et al., 2023).

The proposed framework for LLM-informed visual inspection for infrastructure, which contains three major stages including interactive task planning, command generation and command execution.

In the first stage, the engineer will explain the purpose of the inspection to LLM. LLM will then confirm the information and describe the overall objective and task plan. Through communication and collaboration between LLM and the engineer, a task plan will be established for the next stage. The second stage is command generation, in which LLM will leverage the tools, including sensors, drones and data post-processing models, to fulfil the functions needed in the visual inspection. In this stage the generated command and its explanation will also be the feedback to the engineer to check. After confirmation, the generated command will be executed in the third stage.

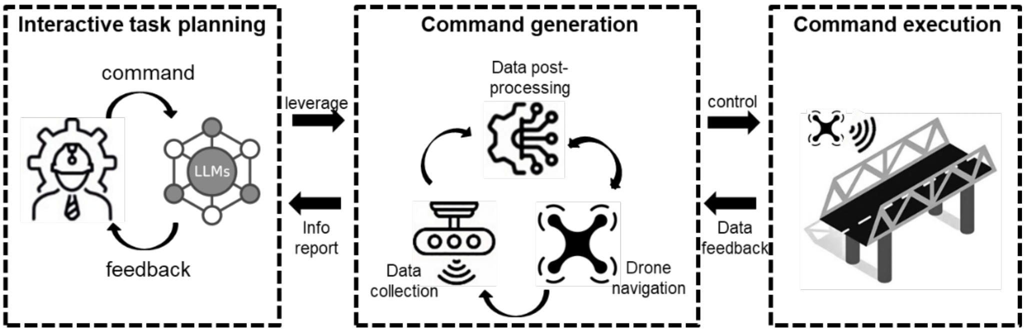

Figure 1: Framework of LLM-informed Drone Visual Inspection for Infrastructure

Specifically, to establish the system, two crucial components are built:

1. High-level function libraries for controlling drones, sensors and post-processing algorithms are built for the efficiency and accuracy of the translating process from human intent to a logical chain of commands.

2. Systematic prompts, including prompt about interacting pattern, navigation and pose planning, prompt about drone control and sensor control, prompt about data post-processing, are designed to facilitate the integration of other task-related AI models or algorithms into a network which is overseen and regulated by the LLM.

In the upcoming section, the two crucial components will be introduced in detail.

3.2 High-level function libraries

High-level function libraries are of great importance for LLM-informed robotics for several reasons:

1. The raw function on the robot platform could not be descriptively enough for LLM to follow, which could influence the efficiency and accuracy of the translating process from human intent to a logical chain of commands.

2. The available functions for LLM should be restricted into some specific region in case for the stability, maintainability, and comprehensibility of the generated code.

High-level function libraries for drone and sensor control should be built specifically to the format or scenario of interest and should map to actual implementations on the robot platform while being named descriptively enough for ChatGPT to follow. The functions of the library are listed in Table 1:

Table 1: Structure of Constructed High-level Function Library

Category	Function	Modules
ControlWrapper	Drone control	Motion control, Attitude control
	Camera control	Recording video, Taking photo, Attitude control, Record position
PostprocessingWrapper	PathPlanner	Path planning, Path restriction
	Data interpretation (image)	3d reconstruction, Image segmentation, Object detection, Save
	Data interpretation (point cloud)	Instance segmentation, Get bounding cuboid, Save

3.3 Systematic prompting construction

In the field of interacting methods between users and LLM (Wang et al., 2023), there are three different ways, including pre-trained language models, prompting and fine-tuning. According to Khot et al. (2023), prompting refers to the interaction methods that focus on calling a model via prompts without involving any parameter updating.

Prompting plays a crucial role in the proposed methodology to define the grounding for LLM in the specific task. There are four major prompting documents in the context of this paper: prompt about the user interface, prompt about rule extraction, prompt about control command of a drone equipped with sensors, and prompt about the working mode of LLM. The relationship between different prompt documents and the corresponding actions of LLM in the validating stage are shown in Figure 2.

Prompt about interacting pattern aims to outline the interaction pattern between the user and LLM, encompassing the Q&A mode and confirmation mechanism. In the Q&A mode, users have the ability to pose questions and receive responses derived from the knowledge stored within the system. Additionally, a confirmation mechanism is in place to validate that the actions carried out by the LLM are in line with the user's intentions.

On the other hand, prompts about navigation and pose planning, prompts about drone control and sensor control, and prompts about data post-processing are created to facilitate LLM's access to the established high-level function libraries, which enables seamless interaction with sensors, drones, various algorithms, and other AI models.

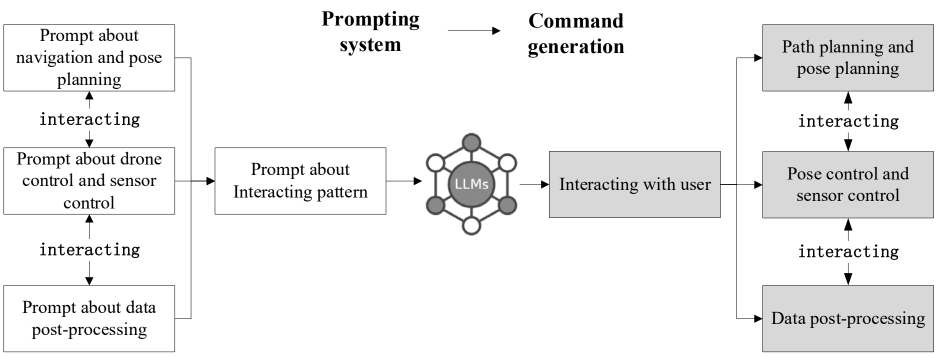

Figure 2: Relationship among Prompts in the Proposed Framework

4. Experiments and discussion

4.1 Simulated experiment

a. Experiment description

A simulated experiment is held to demonstrate the feasibility of the proposed methodology. In this research, we choose Airsim as the platform to build up the simulating environment. Airsim (Shah et al., 2017) is a simulator built on Unreal Engine and includes a physics engine that can operate at a high frequency for real-time hardware-in-the-loop (HITL) simulations with support for popular protocols (e.g., MavLink). It can offer physically and visually realistic simulation, is widely used in the simulation of drone control, and is capable of collecting simulated data such as point clouds and images. In this paper, we first design the simulation field by using maps in the open-source community of Unreal Engine.

As shown in Figure 3, the simulating environment contains one rusty bridge and several buildings to build up a stage for drone-based visual inspection. To simulate defects with different types and degrees, artificial patterns of corrosion and crack are attached on the surface of the components in each simulated structure. It is worth mentioning that the LLM model used in this research is ChatGPT-3.5 for its comprehensively outstanding performance, which has been proven in related research (You et al., 2023).

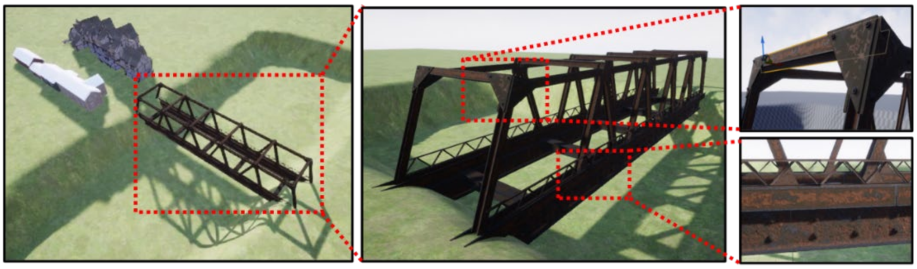

Figure 3: Detail of the Overall Simulating Environment

The high-level function libraries, including drones, sensors, and other data processing models, were built according to the simulating environment to empower the LLM to control the objects. The details of the two libraries are listed in Table 1. On the other hand, the systematic prompts were designed to leverage the LLM's in-context learning capability.

The simulated experiment is designed according to the proposed LLM-informed visual inspection framework. Its detailed description follows.

Table 2: Configuration of the Simulated Sensors

Sensor	Description
GPS	Get position of drone.
RGB-D camera top	Installed on the top of drone: Resolution: 256*192; FOV Degree: 90
RGB-D camera bottom	Installed on the bottom of drone: Resolution: 256*192; FOV Degree: 90 Linked with camera gimbal
Camera gimbal	Set and record pose of RGB-D camera 1.

b. Experiment process and result

In this experiment, we specified two separate tasks:

1. Please inspect component 1 and specify the corrosion region.
2. Please do a 3D reconstruction for component 2 and save the reconstructed point cloud.

Figure 4 shows the interaction process between the user and the LLM agent on task 1. During the interactive task planning stage, the LLM agent firstly specify the target and the objective of the visual inspection, and then separate the objective into a task chain for confirmation.

Based on the task chain, the LLM agent will generate commands and ask engineers to confirm several important issues, including distance from the target and number of control points on the path. This process is shown in Figure 5.

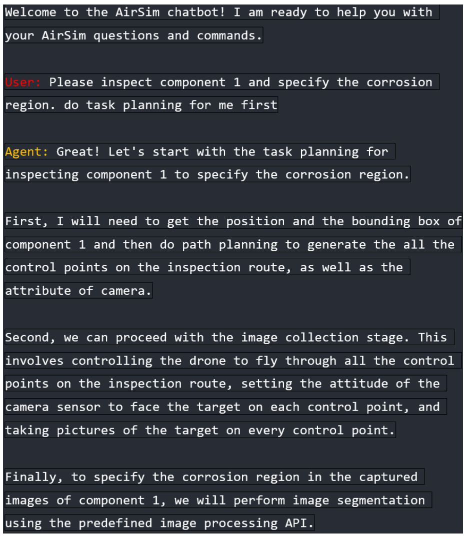

Figure 4: Example of Interacting Dialogue

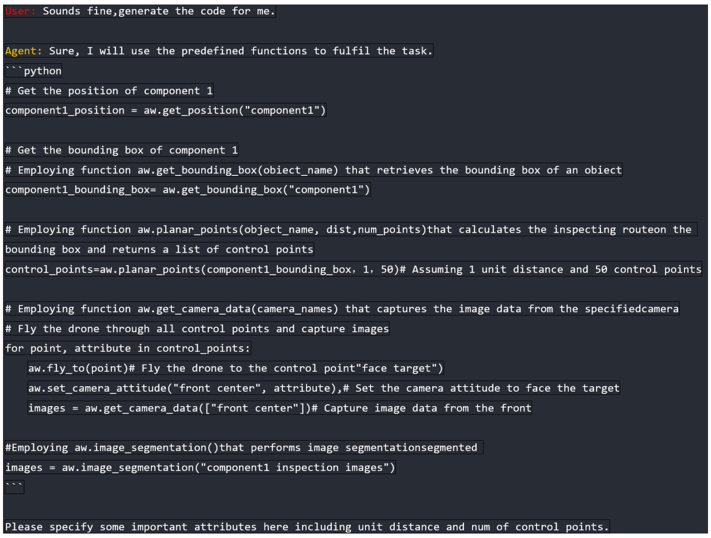

Figure 5: Example of Code Generation

After code generation, code will be executed in the simulated environment to fulfill the task and the result are shown below. Figure 6, Figure 8 and Figure 10 show the key result during the inspection of component 1 on the task of corrosion region specification, while Figure 7, Figure 9 and Figure 11 show the key result during the inspection of component 2 on the task of 3d reconstruction.

Figure 6: Sample of Collected Data (Component 1)

Figure 7: Sample of Collected Data (Component 2)

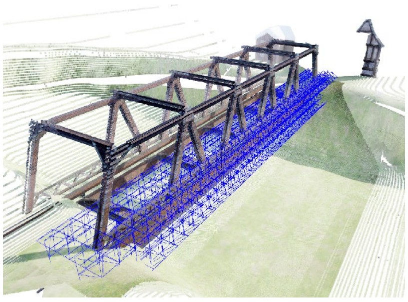

Figure 8: Drone Trajectory (Component 1)

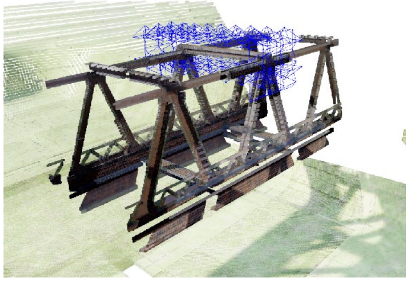

Figure 9: Drone Trajectory (Component 2)

Figure 10: Damage Segmentation Result (Component 1)

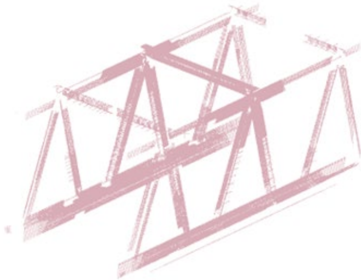

Figure 11: Segmented 3D Point Cloud (Component 2)

4.2 Lab experiment

To prove the effectiveness of the proposed method in the real scenerio, a lab experiment is carried out. Considering the size of the space in the lab and the programming environment that can support the SDK, we chose the Tello EDU, a micro drone developed by DJI, as the lab flight device. It weighs about 80 g and has dimensions of 98 mm * 92.5 mm * 41 mm, respectively. It also comes with a 5MP camera and 2.4 GHz 802.11n Wi-Fi. In this case, a communication connection can be established with the ground equipment through the IP and UDP ports of the drone. This communication link is responsible for passing all relevant commands and information flows, including take-off, landing, forward, backward, left, right, up, down, clockwise rotation, anti-clockwise rotation, taking pictures, etc. In this study, laboratory experiments were conducted on a 160 cm * 160 cm table with four concrete columns. Aruco markers were also placed at the table's four corners and the centre to enhance the 3D reconstruction in the data post-processing session. The detailed lab experimental environment is shown in Figure 12.

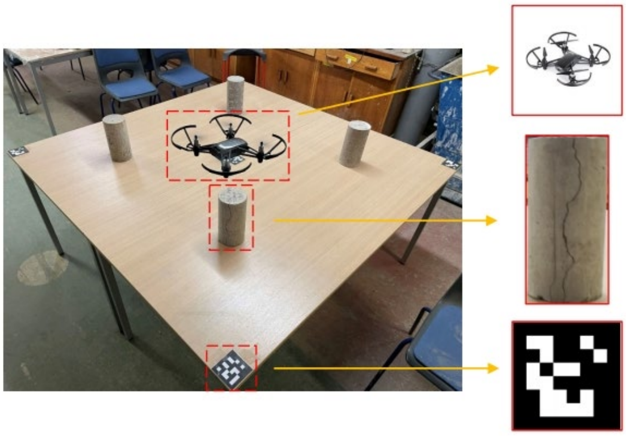

Figure 12: The Lab Experimental Environment

The lab experiment also used the same LLM-driven strategy, and the Tello SDK guide was provided to ChatGPT using previous prompted engineering approach for prior learning of the UAV's flight control commands. However, due to the limitations of the realistic drone hardware equipment, i.e., the lack of GPS and distance sensors, we chose to pre-confirm the world coordinate system position of each object on the table. In this way, the drone can execute tasks along known flight path with the assistance of the LLM. A total of 100 images containing column components were captured during the flight around the table. For these collected data, we performed 3D reconstruction using Agisoft Metashape software providing structure from motion (SFM) algorithm and crack identification using the fine-tuned DeepLabV3+ algorithm. The visual results are shown in Figure 13 and Figure 14.

Figure 13: 3D Reconstruction Result

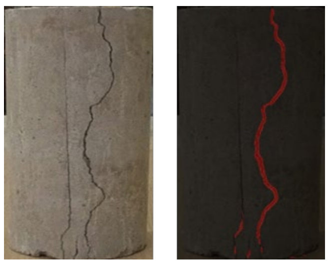

Figure 14: Result of Crack Identification

5. Conclusion

The paper proposes an integrated framework of LLM-informed drone visual inspection for infrastructure to improve the level of automation and interactivity of the inspection process. Additionally, to empower the LLM model to generate code for drone navigation, data acquisition and data post-processing, high-level function libraries about drone control and sensor control are created in this paper. Furthermore, systematic prompts are introduced to explain the proposed inspection scheme, high-level function libraries and engineer-LLM interacting pattern for LLM. The effectiveness of the proposed method is validated in both simulated and laboratory environments.

6. Acknowledgements

This research is funded by the Cardiff University-China Scholarship Council joint program.

References

Ahn, M., Brohan, A., Brown, N., Chebotar, Y., Cortes, O., David, B., Finn, C., Fu, C., Gopalakrishnan, K., Hausman, K., Herzog, A., Ho, D., Hsu, J., Ibarz, J., Ichter, B., Irpan, A., Jang, E., Ruano, R.J., Jeffrey, K., Jesmonth, S., Joshi, N.J., Julian, R., Kalashnikov, D., Kuang, Y., Lee, K.-H., Levine, S., Lu, Y., Luu, L., Parada, C., Pastor, P., Quiambao, J., Rao, K., Rettinghouse, J., Reyes, D., Sermanet, P., Sievers, N., Tan, C., Toshev, A., Vanhoucke, V., Xia, F., Xiao, T., Xu, P., Xu, S., Yan, M., Zeng, A., 2022. Do As I Can, Not As I Say: Grounding Language in Robotic Affordances.

Besada, J., Bergesio, L., Campaña, I., Vaquero-Melchor, D., López-Araquistain, J., Bernardos, A., Casar, J., 2018. Drone Mission Definition and Implementation for Automated Infrastructure Inspection Using Airborne Sensors. Sensors 18, 1170.

Chan, Brodie, Guan, Hong, Jo, Jun, Blumenstein, Michael, 2015. Towards UAV-based bridge inspection systems: a review and an application perspective. Structural Monitoring and Maintenance 2, 283–300.

Chen, B., Xia, F., Ichter, B., Rao, K., Gopalakrishnan, K., Ryoo, M.S., Stone, A., Kappler, D., 2022. Open-vocabulary Queryable Scene Representations for Real World Planning.

Cheng, C., Shang, Z., Shen, Z., 2020. Automatic delamination segmentation for bridge deck based on encoder-decoder deep learning through UAV-based thermography. NDT & E INTERNATIONAL 116.

Lamine, Mohamed, 1 Tazir, Matei Mancas and T Dutoit. “From Words to Flight: Integrating OpenAI ChatGPT with PX4/Gazebo for Natural Language-Based Drone Control.” Proceedings of 2023 the 13th International Workshop on Computer Science and Engineering (2023): n. pag.

Jung, S., Song, S., Youn, P., Myung, H., 2018. Multi-Layer Coverage Path Planner for Autonomous Structural Inspection of High-Rise Structures. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Presented at the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, Madrid, pp. 1–9.

Khot, T., Trivedi, H., Finlayson, M., Fu, Y., Richardson, K., Clark, P., Sabharwal, A., 2023. Decomposed Prompting: A Modular Approach for Solving Complex Tasks.

Nooralishahi, P., Ibarra-Castanedo, C., Deane, S., Lopez, F., Pant, S., Genest, M., Avdelidis, N.P., Maldague, X.P.V., 2021. Drone-Based Non-Destructive Inspection of Industrial Sites: A Review and Case Studies. DRONES 5.

Parliament, E.U., 2011. the Council. Regulation (EU) No 305/2011 of the European Parliament and of the Council of 9 March 2011 laying down harmonised conditions for the marketing of construction products and repealing Council Directive 89/106. EEC.

Shah, D., Osinski, B., Ichter, B., Levine, S., 2022. LM-Nav: Robotic Navigation with Large Pre-Trained Models of Language, Vision, and Action.

Shah, S., Dey, D., Lovett, C., Kapoor, A., 2017. AirSim: High-Fidelity Visual and Physical Simulation for Autonomous Vehicles.

Wang, Z., Zhang, G., Yang, K., Shi, N., Zhou, W., Hao, S., Xiong, G., Li, Y., Sim, M.Y., Chen, X., Zhu, Q., Yang, Z., Nik, A., Liu, Q., Lin, C., Wang, S., Liu, R., Chen, W., Xu, K., Liu, D., Guo, Y., Fu, J., 2023. Interactive Natural Language Processing.

You, H., Ye, Y., Zhou, T., Zhu, Q., Du, J., 2023. Robot-Enabled Construction Assembly with Automated Sequence Planning Based on ChatGPT: RoboGPT. Buildings 13, 1772.

Zhang, C., Zou, Y., Wang, F., del Rey Castillo, E., Dimyadi, J., Chen, L., 2022. Towards fully automated unmanned aerial vehicle-enabled bridge inspection: Where are we at? Construction and Building Materials 347, 128543.